MACHINE TRANSLATION SYSTEM WITH IDIOM REGISTERING FUNCTION

Publication number: JP8055123

Publication date: 1996-02-27

Inventor: OKUNISHI TOSHIYUKI; FUKUMOCHI YOUJI; SADA ICHIKO; KUTSUMI TAKESHI

Applicant: SHARP KK

Classification:

- international: ***G06F17/27; G06F17/28;*** **G06F17/27; G06F17/28;** (IPC1-7): G06F17/28; G06F17/27

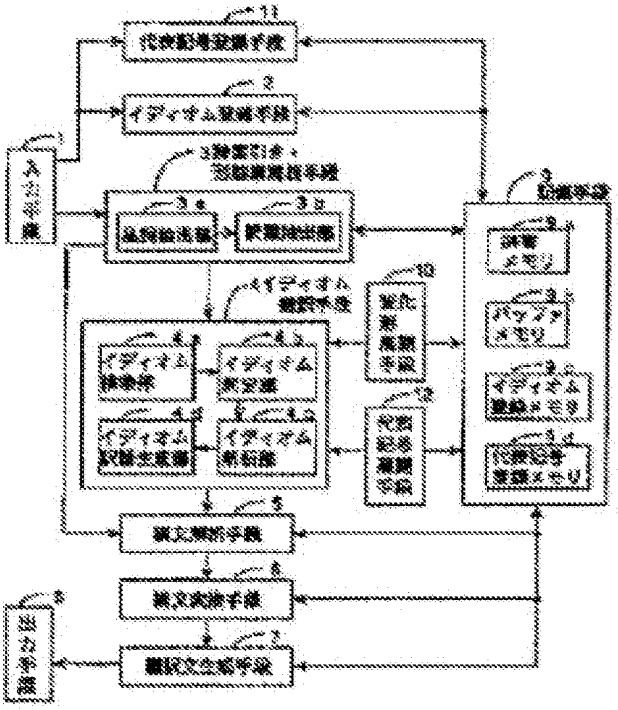
- European:

Application number: JP19940186127 19940808

Priority number(s): JP19940186127 19940808

Abstract of **JP8055123**

PURPOSE:To provide a machine translation system capable of corresponding also to a modification form of a fixed part in an index word of an idiom and leading a user definition mark into a variable part in respect to a machine translation system with an idiom registering function. CONSTITUTION:The machine translation system is provided with an idiom registering means 2 for registering an entry word and a translated word of an idiom so that the fixed part of the idiom is expressed by a normal word, a normal word string or a modification developing mark representing the modified expression of the word or the word string and the variable part of the idiom is expressed by the composite format of a 1st representative mark representing the set of words or word strings sharing a prescribed attribute, a modification form developing means 10 for generating and developing the fixed part of the index word of an idiom to be translated to all previously set modified expressions and an idiom translation means 4 for identifying an input character string or a part of the character string with the index word or the like obtained by developing the fixed part of the index word of the idiom is developed to a modified expression.



(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平8－55123

(43)公開日　平成８年(1996)２月27日

(51)Int.Cl.⁶	識別記号	庁内整理番号	Ｆ Ｉ	技術表示箇所
Ｇ０６Ｆ　17/28				
17/27				
		8420－5L	Ｇ０６Ｆ　15/ 38	Ｃ
		8420－5L		Ｅ

審査請求　未請求　請求項の数３　ＯＬ　（全 16 頁）

(21)出願番号　特願平6－186127

(22)出願日　平成６年(1994)８月８日

(71)出願人　000005049
シャープ株式会社
大阪府大阪市阿倍野区長池町22番22号

(72)発明者　奥西　稔幸
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内

(72)発明者　福持　陽士
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内

(72)発明者　佐田　いち子
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内

(74)代理人　弁理士　野河　信太郎

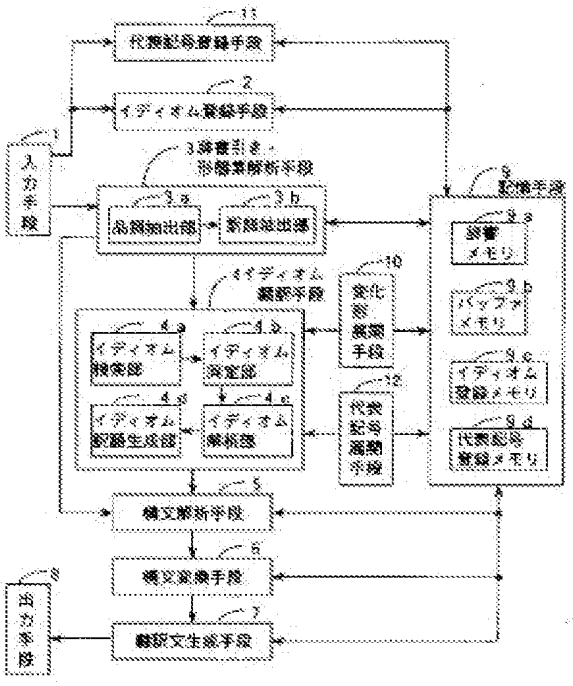
最終頁に続く

(54)【発明の名称】　イディオム登録機能を有する機械翻訳装置

(57)【要約】

【目的】　イディオム登録機能を有する機械翻訳装置に関し、イディオムの見出し語の固定部分の変化形にも対応でき、可変部分に利用者定義の記号を導入することができる機械翻訳装置を提供することを目的とする。

【構成】　イディオムの固定部分が通常の単語、単語列、又はその単語もしくは単語列の変形表現を代表する変化展開記号によって表現され、かつイディオムの可変部分が所定の属性を共有する単語又は単語列の集合を代表する第１の代表記号を複合した形式で表現される，見出し語とそのイディオムの訳語を登録するイディオム登録手段と，翻訳すべきイディオムの見出し語に対してその固定部分を予め設定されたすべての変形表現に生成展開する変化形展開手段と、入力文字列あるいはその一部分と、イディオムの見出し語の固定部分が変形表現に展開された見出し語等との同定を行うイディオム翻訳手段とを備えたことを特徴とする。



1

【特許請求の範囲】

【請求項1】　文字列および記号を入力する入力手段と、

予め定められた単語又は単語列からなる固定部分と、共通の属性を持つ単語又は単語列に変化可能な可変部分とからなるイディオムに対して、前記固定部分が通常の単語、単語列、又はその単語もしくは単語列の変形表現を代表する変化展開記号によって表現され、かつ前記可変部分が所定の属性を共有する単語又は単語列の集合を代表する第1の代表記号を複合した形式で表現される見出し語とそのイディオムの訳語を登録するイディオム登録手段と、

イディオムの登録と翻訳処理に必要な辞書及び処理結果を記憶する記憶手段と、

入力単語列を形態素に分解し、かつ文法解析を行う辞書引き・形態素解析手段と、

翻訳すべきイディオムの見出し語に対してその固定部分を予め設定されたすべての変形表現に生成展開する変化形展開手段と、

入力文字列あるいはその一部分と、登録されたイディオムの見出し語あるいは前記変化形展開手段によってその見出し語の固定部分が変形表現に展開された見出し語との同定を行い、同定されたイディオムの見出し語に対応する文字列の訳語を生成するイディオム翻訳手段と、

構文解析手段と、構文変換手段と、翻訳文生成手段と、翻訳文を出力する出力手段とを備えたことを特徴とするイディオム登録機能を有する機械翻訳装置。

【請求項2】　前記変化形展開手段が、イディオムの見出し語の固定部分を、固定部分を構成する単語を活用変化させた表現形式又はその固定部分に助動詞もしくは否定副詞を連接させた表現形式に生成展開することを特徴とする請求項1のイディオム登録機能を有する機械翻訳装置。

【請求項3】　新たに定義された属性とその属性値を有する単語又は単語列を代表する第2の代表記号を前記記憶手段に登録する代表記号登録手段と、

翻訳すべきイディオムの見出し語の可変部分に含まれる前記第2の代表記号を定義された属性とその属性値とに生成展開する代表記号展開手段とを備え、

前記イディオム登録手段が、前記入力手段によって入力された第1の代表記号および／または第2の代表記号を用いて表現されるイディオムの見出し語とその訳語を登録し、前記イディオム翻訳手段が、入力文字列あるいはその一部分の属性及びその属性値と、前記代表記号展開手段によって生成展開された見出し語の第2の代表記号の属性及びその属性値との同定を行うことを特徴とする請求項1または2記載のイディオム登録機能を有する機械翻訳装置。

【発明の詳細な説明】

【0001】

2

【産業上の利用分野】この発明は、電子化辞書あるいは電子化辞書を搭載した情報検索装置あるいは電子化辞書を搭載した機械翻訳装置に関し、特に、可変部分を含むイディオムを登録し検索・翻訳することのできるイディオム登録機能を有する辞書検索装置に関する。

【0002】

【従来の技術】現在実用化されている言語処理装置には、人間の文書作成活動を支援するためのワードプロセッサや、ある言語で書かれた文書を別の言語に翻訳するための機械翻訳装置などがある。これらの言語処理装置には、それぞれの目的に応じた情報を納めた辞書が備えられている。ここでいう辞書とは、見出語とそれに付帯する各種の情報の組とを1単位の項目としたものを多数統合し、見出語を用いて所望の項目を容易に検索できるように系統立てて並べたものである。

【0003】辞書は、原則として機械可読な不揮発性の媒体に機械可読な形式で記録される。このような辞書を、電子化辞書と呼ぶことにする。電子化辞書を機械翻訳において用いる場合には、見出語としては原語（ソース原語）の単語列（1単語のみのものも含む）が用いられ、その単語列に付帯する各種の情報として、見出語の品詞、形態属性、訳語、訳語の品詞等の情報が用いられる。

【0004】このような言語処理装置を用いて利用者が処理あるいは作成しようとしている文書に、この装置に備えられた辞書に見出語として記載されていない単語が含まれている場合には、作業効率が著しく低下してしまう。そのために、辞書に収録する見出語は、より多いほうが好ましい。また、機械翻訳の場合には、原語の各単語のみではなく、イディオムを見出語として採用し、対応するターゲット言語の言い回し等をペアとして、このようなペアをできるだけ多数登録しておくことが翻訳効率の上では望ましい。

【0005】通常イディオムには、数詞、所有格代名詞、再帰代名詞など、主語や他の語との関係においてその形を変えるイディオムが多い。例えば[do one's best]中のone's は主語に応じてyour、my、his、herなどの所有格代名詞となる。翻訳処理並びに辞書開発の効率上、このようなイディオムは、具体的な語を入れたイディオムを全て列挙するのではなく、見出し語の一部分の単語としてある文法特徴を共有する単語や句ならば任意の単語が入るような形で登録できるのが好ましい。以下では、ある文法特徴を共有する単語や句ならば任意の単語が入る部分を可変部分と呼び、それ以外のイディオムの骨格となる単語や単語列の部分を固定部分と呼ぶ。

【0006】また、1つのイディオムには、複数の可変部分が存在するものがあり、可変部分には単語だけでなく名詞句や文が適用される場合もある。このような種々の可変部分を表現するために、＊に続く記号（以下この記号を代表記号と呼ぶ）を導入する機械翻訳装置が提案

3

されている。

【0007】たとえば「～よりN倍…」という日本語訳をもつ英単語イディオム“N timesas…as～”を登録する場合、次のように記述することができる。

英単語　〔 *m times as *ad as *CN 〕
訳語　*CN より　*m　倍　*ad

見出し（英単語列）の中で、先頭に“*”のついた単語（*m, *ad, *CN）、すなわち代表記号が可変部分であり、それ以外の単語（times,as）が固定部分である。

【0008】可変部分では、代表記号で表す品詞（上例では、m:数詞、ad: 形容詞、C:文、N:名詞句、CN:CまたはN）の任意の単語と適合できるのに対し、固定部分ではその表記を持つ単語とでないと適合できない。なお、可変部分に指定できる代表記号は予めシステムに定義されたものである。

4

＊【0009】今、

〔 This apple is three times as big as that orange. 〕

という英文が入力されたとし、以下に機械翻訳処理の概要を以下に示す。まず、各単語の辞書引きが行われる。その結果、次のような単語情報が得られる。

three　数詞(m)
big　形容詞(ad)
that　冠詞(d)
orange　名詞(n)

【0010】次に、イディオムの検索と解析処理が実行され、固定部分ならびに可変部分のマッチングが行われる。このとき、入力文の一部が上記のように登録されたイディオムに適合することがわかる。

＊

〔 This apple is three times as big as that orange. 〕
　　*m　#### ## *ad ##　*CN

ここで、可変部分*mと“three”がマッチングし、*adと“big”がマッチングし、*CNと“that orange”がマッチングする。また、“times as”と“as”がイディオムの固定部分（#記号部）としてマッチングする。

【0011】次に、イディオムの可変部分の訳を生成する。

※CN　あのオレンジ
*m　3
*ad　大きい

最後に、固定部分の訳と合成することで入力文に対する日本語文を生成する。

※

「この　りんごは　あのオレンジ　より　3　倍　大きい」。
　　　　　　　　　####　##

【0012】

【発明が解決しようとする課題】しかしながら、以上に示した機械翻訳システムでは、

（1）イディオム見出しの固定部分に、活用変化が書けない。

（2）イディオム見出しの可変部分に、システムで定義された代表記号以外の記号が使えない。

という2つの制限がある。このため、イディオムに登録する見出し語が増大し、その結果記憶容量及び検索時間が増大するという問題や、利用者、すなわち辞書にイディオムを登録する者の負担が大きくなるという問題が発生する。

【0013】固定部分と可変部分から成るイディオム見出し語の固定部分と入力文のマッチングは文字列だけで比較するので、単複変化する名詞、活用変化する動詞、助動詞、形容詞、また、助動詞や否定副詞(not)が付加する単語を含む入力文の場合は、通常よく使われる基本形のイディオムを登録するだけではマッチングが失敗することとなる。そのため、以上に示したような変化形を全て異なる見出しとして列挙する必要があった。

【0014】『～することはできない』という日本語訳を持つ英単語イディオム“There isno ～ing.”を例に説明する。このイディオムは、次のような形式で登録される。

英単語　〔There is no *Ving〕
訳語　*Ving ことができない。

ここで、“*Ving”は『動詞の進行形』を表す代表記号である。

【0015】しかし、“There was no going to school yesterday.”という文はこのイディオムを含むにも関わらず過去形の文であるため、文中の“There was no”が見出しの固定部分の“There is no”と一致できない。すなわち、一部に動詞等の語形変化する単語を含んだイディオムの場合、変化形を全て展開して登録する必要があり、この例では〔There was no〕や〔There will be no〕という見出しのイディオムを別に登録する必要があった。

【0016】固定部分と可変部分から成るイディオム見出しの可変部分は、あらかじめ機械翻訳システムで定義されたもののみが使用される。これは、幅広い利用者がターゲットである商用化システムにおいて、利用者のあらゆる要求を予測し代表記号を全て準備しておくのは現実には不可能であるし、非常に特殊なものまでシステムで定義しておくのも効率上問題があったからである。

【0017】このような従来の機械翻訳システムにおいて、所望する代表記号が定義されていなかった場合について、『できる限り～』という日本語訳を持つ英単語イディオム“as～as…can”を例に説明する。2つのasの間の『～』に入れることができる語句は、形容詞原級か、形容詞原級で修飾される名詞句であるが、もし、シ

5

ステムに定義された代表記号に「形容詞を含む名詞句」という記号がなかったとすると，利用者はシステムが定義している代表記号で最も近いもの，例えば，次のように，*3（名詞句）とだけ指定しているイディオムを使用するしかない。

英単語　[as *N as *a *3 can]

訳語　できる限り *3

ここで，*3は主格代名詞を表す代表記号である。

【００１８】しかし，このような近似的な登録では実際＊

[I regard the person as policeman as you can see.]

**　　*N　　**　*3 ***

というように，policeman が*Nとマッチしてしまい，この文がイディオムを含んでいると誤って認識してしまう。

【００１９】さらに，イディオムと解釈したas policeman as you can が主語でsee が動詞からなる関係代名詞の先行詞がthe personであるという誤った解析がたまたま成立し，結果的に次のような誤訳を生成してしまう。

「私はできる限り警官が見る人を考慮する。」

【００２０】このようにシステムで定義された代表記号しか使えないため，近似的な指定しかできず，利用者が意図しなかった誤訳が生じる場合もある。また，指定を詳細にすると見出しが複雑になる場合が多い。このことは，類似した詳細な条件が複数現れた場合に特に顕著になる。このように，従来の機械翻訳装置では，以上述べたような２つの制限があるため，イディオムに登録する見出し語が増大しその結果記憶容量及び検索時間が増大するという問題や，利用者の負担が大きくなるという問題が発生していた。

【００２１】そこで，この発明は，以上のような事情を考慮してなされたものであり，イディオムの見出し語の固定部分の変化形にも対応でき，可変部分に利用者定義の記号を導入することができるイディオム登録機能を持つ機械翻訳装置を提供することを目的とする。

【００２２】

【課題を解決するための手段】図１に，この発明の基本構成ブロック図を示す。同図において，この発明は文字列および記号を入力する入力手段１と，予め定められた単語又は単語列からなる固定部分と，共通の属性を持つ単語又は単語列に変化可能な可変部分とからなるイディオムに対して，前記固定部分が通常の単語，単語列，又はその単語もしくは単語列の変形表現を代表する変化展開記号によって表現され，かつ前記可変部分が所定の属性を共有する単語又は単語列の集合を代表する第１の代表記号を複合した形式で表現される見出し語とそのイディオムの訳語を登録するイディオム登録手段２と，イディオムの登録と翻訳処理に必要な辞書及び処理結果を記憶する記憶手段９と，入力単語列を形態素に分解し，かつ文法解析を行う辞書引き・形態素解析手段３と，翻訳すべきイディオムの見出し語に対してその固定部分を予

6

＊にはイディオムを含んでいない文でも誤って認識して，翻訳精度の低下を招く場合がある。例えば，

[I regard the person as policeman as you can see.]

あなたがわかるので，私はその人を警官とみなす。

という入力文は，２つのasの間の名詞句が形容詞を含んでいないため本来ならば上記イディオムの文ではないが，上記登録では

め設定されたすべての変形表現に生成展開する変化形展開手段１０と，入力文字列あるいはその一部分と，登録されたイディオムの見出し語あるいは前記変化形展開手段１０によってその見出し語の固定部分が変形表現に展開された見出し語との同定を行い，同定されたイディオムの見出し語に対応する文字列の訳語を生成するイディオム翻訳手段４と，構文解析手段５と，構文変換手段６と，翻訳文生成手段７と，翻訳文を出力する出力手段８とを備えたことを特徴とするイディオム登録機能を有する機械翻訳装置を提供するものである。

【００２３】また，前記変化形展開手段１０が，イディオムの見出し語の固定部分を，固定部分を構成する単語を活用変化させた表現形式又はその固定部分に助動詞もしくは否定副詞を連接させた表現形式に生成展開するようにしてもよい。

【００２４】また，新たに定義された属性とその属性値を有する単語又は単語列を代表する第２の代表記号を前記記憶手段９に登録する代表記号登録手段１１と，翻訳すべきイディオムの見出し語の可変部分に含まれる前記第２の代表記号を定義された属性とその属性値とに生成展開する代表記号展開手段１２とを備え，前記イディオム登録手段２が，前記入力手段１によって入力された第１の代表記号および／または第２の代表記号を用いて表現されるイディオムの見出し語とその訳語を登録し，前記イディオム翻訳手段４が，入力文字列あるいはその一部分の属性及びその属性値と，前記代表記号展開手段１２によって生成展開された見出し語の第２の代表記号の属性及びその属性値との同定を行うようにしてもよい。

【００２５】また，前記記憶手段９は，入力された文字列の翻訳を行うための文法および訳語情報を持つ辞書メモリ９ａと，訳語生成に至るまでの処理の結果を記憶するバッファメモリ９ｂと，前記イディオム登録手段２によって登録されたイディオムを記憶するイディオム登録メモリ９ｃとから構成することが好ましい。

【００２６】前記辞書引き・形態素解析手段３は，入力された文字列を単語に分解し各単語の品詞情報を生成する品詞抽出部３ａと，各単語の訳語の候補を生成する訳語抽出部３ｂとから構成することが好ましい。

【００２７】前記イディオム翻訳手段４は，前記イディ

7

オム登録メモリを検索し分解された単語列と表現形式が一致可能なイディオムの見出語の候補を選択するイディオム検索部4ａと，イディオムの中の代表記号の位置に相当する単語又は単語列の属性が，代表記号に与えられた属性に一致するイディオムの見出語をイディオム候補の中から一つに特定するイディオム同定部4ｂと，代表記号に対応する単語又は単語列の構文を解析しイディオム全体の文構成を生成するイディオム解析部4ｃと，イディオムの文構成を基に，入力された単語列のイディオム部の訳語を生成するイディオム訳語生成部4ｄとから構成することが好ましい。

【００２８】ここで，図1において入力手段1としては，キーボード，又はマウス，ペンあるいはトラックボールなどのポインティングデバイスが用いられるがこれに限定されるものではなく，その他の入力装置を用いてもよい。また，記憶手段9は通常ＲＯＭ，ＲＡＭ，フロッピーディスク又はハードディスク等が用いられるが，これに限定されるものではなく，その他の記憶装置を用いてもよい。特に，辞書メモリ9ａはＲＯＭが好ましく，バッファメモリ9ｂ，イディオム登録メモリ9ｃ及び代表記憶メモリ9ｄはＲＡＭが好ましい。

【００２９】また，イディオム登録手段2，代表記号登録手段11，辞書引き・形態素解析手段3，イディオム翻訳手段4，構文解析手段5，構文変換手段6及び翻訳文生成手段7としては，通常ＣＰＵが用いられ，ＲＯＭ，ＲＡＭ，Ｉ／Ｏインターフェイス等の周辺回路を含んだマイクロコンピュータを用い，ＲＯＭ又はＲＡＭにはこの文書処理装置の動作を制御するプログラムが内蔵されていることが好ましい。

【００３０】

【作用】イディオムの見出し語および訳語を登録する場合，イディオムの固定部分は通常の単語，単語列又はその単語もしくは単語列の変形表現を代表する変化展開記号によって表現され，イディオムの可変部分は所定の属性を共有する単語又は単語列の集合を代表する第1の代表記号を複合した形式で表現されて，イディオム登録手段2が記憶手段9に登録する。

【００３１】これにより，可変部分の属性が共通し，さらに固定部分が種々の変形表現されることのある複数個のイディオムの見出し語を1つの見出し語で登録することができる。すなわち，この発明によれば，イディオムの見出し語としてその固定部分及び可変部分に対して考えられるあらゆるパターンを登録する必要はなく，登録されるイディオムの見出し語の増大を抑えることができる。

【００３２】入力された文字列を単語に分解し，分解された単語列の一部の表現形式と一致可能なイディオムの見出し語の候補を検索し，さらにその見出し語の固定部分の中に変化展開記号がある場合には，変化形展開手段によって生成展開された変形表現とその変化展開記号が

8

存在する位置に相当する単語又は単語列との同定を行い，イディオムを特定する。

【００３３】以上のように，この発明によれば，入力文のうちあるイディオムの固定部分に相当する単語又は単語列が，そのイディオムの見出し語として登録されている単語又は単語列とは完全に一致しないがその変形表現と一致する場合にも，イディオムの同定をすることができる。

【００３４】また，イディオムの可変部分は新たに定義された属性とその属性値を有する単語又は単語列を代表する第2の代表記号を含む形式で表現されて記憶手段9に登録される。

【００３５】そして，入力された文字列を単語に分解し，分解された単語列の一部の表現形式と一致可能なイディオムの見出し語の候補を検索し，さらにその見出し語の可変部分の中に第2の代表記号がある場合には，代表記号展開手段によって生成展開された属性及びその属性値と，その第2の代表記号が存在する位置に相当する単語又は単語列の属性及びその属性値との同定を行い，イディオムを特定する。

【００３６】以上のように，この発明によれば，新たに定義された属性及びその属性値を有する単語又は単語列を代表する第2の代表記号を用いてイディオムの見出し語の可変部分を表現し，かつイディオムの同定を行うので，イディオムとして登録する見出し語の長さを抑えることができると共に，利用者自身によって必要な，あるいは，ある分野の文章に特有な表現形式を持つイディオムの登録及び同定をすることができる。

【００３７】

【実施例】以下，図に示す実施例に基づいて，この発明を詳述する。なお，これによってこの発明が限定されるものではない。実施例の説明の前に，機械翻訳の概念について簡単に説明する。図2を参照して，機械翻訳において行なわれる解析処理には，様々な解析レベルがある。機械翻訳は，図2の左上に示されるソース言語が入力された場合に，各レベルの処理を順に行なって最終的に図2の右側に示されるターゲット言語を得るための処理である。すなわちソース言語が入力されると，まずレベルＬ1の辞書引き処理，レベルＬ2の形態素解析処理，レベルＬ3の構文解析処理，…と処理が進められ，最終的にレベルＬ10の形態素生成処理が行なわれてターゲット言語が生成される。

【００３８】機械翻訳は，どのレベルの解析処理まで行なうかによって，大きく次の2つに分けられる。第1は，レベルＬ6に示されるソース言語およびターゲット言語のどちらにも依存しない概念である中間言語まで解析し，そこからレベルＬ7の文脈生成，レベルＬ8の意味生成，レベルＬ9の構文生成，レベルＬ10の形態素生成へと進み，ターゲット言語を生成していくピボット方式である。第2は，上述のレベルＬ2の形態素解析，

9

レベルL３の構文解析，レベルL４の意味解析およびレベルL５の文脈解析のいずれかまで解析を行なってソース言語の内部構造を得，次に，得られたソース言語の内部構造と同じレベルのターゲット言語の内部構造に変換した後，ターゲット言語を生成するトランスファー方式である。

【００３９】以下，図２に示される各解析処理の内容について説明する。

（１）　辞書引き，形態素解析

ここでは，形態素が格納された辞書を参照しながら入力された文章を形態素列（単語列）に分割し，この各単語に対する品詞などの文法情報および訳語を得，さらに時制・人称・数などを解析する処理が行なわれる。

【００４０】（２）　構文解析

ここでは，単語間の係り受けなどの文章の構造（構造解析木）を決定する処理が行なわれる。

（３）　意味解析

複数の構造解析の結果から，意味的に正しいものとそうでないものとを判別する処理が行なわれる。

（４）　文脈解析

文脈解析処理では，入力された文章の話題を理解し，入力文章中に含まれる省略部分や曖昧さなどを取去る処理が行なわれる。

【００４１】次に，図３に示すこの発明の一実施例である機械翻訳装置のブロック図について説明する。同図において，３１はメインＣＰＵ（中央処理装置），３２はメインメモリ，３３はＣＲＴ（陰極線管）やＬＣＤ（液晶表示装置）などからなる表示装置，３４はキーボード，３５は翻訳モジュール，３６は翻訳モジュール３５に接続された翻訳用の辞書，文法規則および木変換構造規則などを格納している辞書メモリ，３７は上記構成部品を接続するバスである。

【００４２】また，辞書メモリ３６には，イディオムや，利用者が独自に定義した代表記号を格納しておくことのできる記憶領域を備える。ＣＰＵ３１は，イディオムの登録及び代表記号の登録の処理と，後述する翻訳モジュール３５の処理の制御を行う。

【００４３】翻訳モジュール３５は，ソース言語の文章が入力されると，それを所定の手順で翻訳してターゲット言語を出力するものである。すなわち，キーボード３４から入力されたソース言語はメインＣＰＵ３１の制御により翻訳モジュール３５に送られる。翻訳モジュール３５は辞書メモリ３６に記憶されている辞書，文法規則および木構造変換規則等を用いて，入力されたソース言語を後に詳述するようにしてターゲット言語に翻訳する。その結果は，メインメモリ３２に一旦記憶されると共に，表示装置３３に表示される。

【００４４】図４に翻訳モジュール３５のブロック図を示す。翻訳モジュール３５は，バス３７に接続され，バス３７を介して入力されるソース言語を，所定の翻訳プログラムに従って翻訳してターゲット言語としてバス３７に出力するための翻訳ＣＰＵ４５と，バス３７に接続され，翻訳ＣＰＵ４５で実行される翻訳プログラムを格納するための翻訳プログラムメモリ４６と，入力されたソース言語の原文を各単語ごとに格納するためのバッファＡ（４０）と，バッファＡ（４０）に格納された各単語につき，辞書メモリ３６に含まれる辞書を参照して得た各単語の品詞，訳語などの情報を格納するためのバッファＢ（４１）と，ソース言語の構造解析木に関する情報を格納するためのバッファＣ（４２）と，ソース言語の構造解析木から変換されたターゲット言語の構造解析木を格納するためのバッファＤ（４３）と，バッファＤ（４３）に格納されたターゲット言語の構造解析木に適切な付属語（日本語ならば助詞や助動詞など）を補充して，ターゲット言語の形として整えられた文章を格納するためのバッファＥ（４４）とを含む。

10

【００４５】以上のような構成を持つ翻訳モジュール３５において，少なくとも図２に示したレベルL３の構文解析のレベルまでの解析を行うものとする。ここで，翻訳処理手順を記述した前記翻訳プログラムは，辞書引き・形態素解析部，イディオム翻訳部，構文解析部，構文変換部，翻訳文生成部，変化形展開部及び代表記号展開部から構成される。

【００４６】以下，図３～図１０を参照して，本実施例の機械翻訳装置による英日翻訳の動作を説明する。ここでは，イディオムを含まない英文“This is a pen.”を例にとって，この英文を日本文に翻訳する動作の概要を示す。

【００４７】まず，読込まれた原文は形態素解析によって形態素に分解され，図５に示されるようにバッファＡ（４０）（図４参照）に格納される。続いて翻訳プログラムメモリ４６に記憶されたプログラムに基づく翻訳ＣＰＵ４５の制御の下に，辞書引き・形態素解析部によって，バッファＡ（４０）に格納された原文の各単語ごとに，辞書メモリ３６に格納されている辞書を参照することにより各単語の訳語などの情報が得られる。たとえば，その情報の一部である品詞情報は，図６のようにバッファＢ（４１）に格納される。

【００４８】ここで，“this”の多品詞語であって代名詞，指示形容詞の２つの品詞を持つ，また“is”の品詞は動詞である。同様に“a”，“pen”についてもそれぞれの品詞がバッファＢ（４１）に格納される。“this”は多品詞語であるが，文中の品詞が何であるかについては，翻訳プログラムのうち構文解析部に相当する処理によって一意に決定される。

【００４９】翻訳プログラムのうち構文解析部に相当する処理においては，辞書メモリ３６に格納された辞書および文法規則に従って，各単語間の係り受け関係を示す構造解析木がたとえば図７に示されるように決定される。この構文解析結果は図４のバッファＣ（４２）に格

11

納される。

【００５０】構造解析木の決定は次のようにして行なわれる。辞書メモリ３６に格納された文法規則から、英語に関する文法規則として次のようなものが得られる。

文→主部，述部

主部→名詞句

述部→動詞，名詞句

名詞句→代名詞

名詞句→冠詞，名詞

【００５１】この規則のうちたとえば１つ目の規則は、「文は主部と述部からできている。」ということを表わす。以下，これらの規則に従って構造解析木が決定される。なお、このような文法規則は同じように日本語についても用意されており、英語の文法規則と日本語の文法規則との間で対応づけがなされている。

【００５２】翻訳プログラムのうち、構文変換部に相当する処理においては、辞書メモリ３６の木構造変換規則を用いて、入力された英文の構造解析木（図７参照）の構造が、図８に示される日本文に対する構文解析木の構造に変換される。得られた結果は図４に示されるバッファＤ（４３）に格納される。この説明において用いられている例文"This is a pen."は、この変換によって日本語文字列「これ　ペン　である」に変換されたことになる。

【００５３】翻訳プログラムのうち翻訳文生成部に相当する処理を行なう部分は、得られた日本語文字列「これ　ペン　である」に適切な助詞「は」や助動詞をつけることにより、図９に示されるような日本語の形にし、図４のバッファＥ（４４）に格納する。この得られた日本文「これはペンである。」は、図３に示される翻訳モジュール３５から出力され、メインメモリ３２に格納されるとともに、表示装置３３に表示される。

【００５４】以上が、イディオムを含まない文の翻訳処理の概要であるが、イディオムを含む文の翻訳処理においては、上記処理のほか、イディオム翻訳部におけるイディオムの同定、解析及び訳語の生成処理が行われ、さらに、イディオム翻訳部の処理に関連して、変化形展開部及び代表記号展開部の処理が行われる。

【００５５】ここで変化形展開部は、後述するように、イディオムの固定部分に対して活用変化などの変化形を考慮したマッチング処理を行うものである。また、代表記号展開部は、利用者が独自に定義した代表記号に対してマッチング処理を行うものである。以上の各部の処理は、翻訳モジュール３５の翻訳ＣＰＵ４５によって翻訳プログラムの手順に従って実行される。

【００５６】次に、図１０～１４を用いて、見出し語のうち固定部分を変化形展開記号で表現したイディオムの登録について説明する。変化形展開記号は、次のような記号である。

＊品詞記号（単語）

12

なお、「品詞記号」は、変化対象の「単語」が多品詞だった場合に、どの品詞で変化されるかを指定するためのものである。

【００５７】たとえば、"as～as can be"というイディオムは次のように記述される。

英単語［ as *a as *x(can) be. ］

訳語　この上なく*aで

ｘは助動詞を表す品詞記号であり、助動詞としてのcanの過去形couldであっても、このイディオムであることを表している。品詞記号ｘを指定するのは、canが多品詞語だからである。

【００５８】すなわち，canは名詞又は動詞としての用法もあり、品詞の指定がないと、名詞として変化(cansなど)させるべきか、又は動詞として変化(canned)させるべきかを特定するのが困難だからである。

【００５９】また、ここで最後のbeには変化展開指定がない。これは、canまたはcouldのどちらかであろうが，be動詞は原形しかありえないからである。このように、変化形を持つが見出し以外の形は認識してはいけない時には、変化展開指定をしない。なお変化させるイディオム登録の方が多い場合には、変化指定のデフォルトを逆にして、無変化の場合に指定させてもよい。

【００６０】以下に、

［ I was as happy as could be.］

という入力文があった場合を例にとり、このイディオムを使った翻訳内容を説明する。図１０に，イディオム部分の翻訳処理のフローチャートを示す。

【００６１】まず、図１０のステップＳ１１１～Ｓ１１２において、入力文の先頭単語から順次辞書引きが行われる。代表記号を含んだイディオムも他の単語と同様に基本辞書に登録されているので、３単語目（ｓ＝3の時）asの辞書引き中に，［as *aas *x(can) be］の見出しが検索される（ステップＳ１１３、Ｓ１１７、Ｓ１１８）。

【００６２】次に、ステップＳ１１４において、イディオムの見出し中の各単語と入力文の単語の間でマッチングが行われる。イディオム中の固定部分の単語と入力文の単語の間のマッチングは文字列比較だけで高速に処理できるので，最初に固定部分だけのマッチング処理を行う。

【００６３】ここで，この例では，

［ as *a as *x(can) be. ］と" as happy as could be."

の間での固定部分のマッチング処理に入る。

【００６４】図１１に、この固定部分のマッチング処理のフローチャートを示す。まず、イディオム中の単語番号を示す変数ｐを初期化する（ステップＳ１２１）。ステップＳ１２２において、イディオム中の単語番号ｐの単語Ｗｐが固定部分であるかどうか判断し、固定部分でない，すなわち可変部分である場合は，ステップＳ１３

15

2、Ｓ１３３へ処理を進め、すべての単語が調べられるまで、処理を繰り返す。

【００６５】また、単語W_iが固定部分である場合は、ステップＳ１２３へ進み、単語W_iが変化形展開指定かどうか判断する。単語W_iが変化形展開指定の場合は、ステップＳ１２５、Ｓ１２６へ進み、活用変化させたマッチング処理に入るが、単語W_iが変化形展開指定でない場合は、単語W_iとW_jとの比較を行い（ステップＳ１２４）、一致する場合は、次の単語に対するマッチング処理を繰り返す（ステップＳ１３０～１３３、Ｓ１３２）。

【００６６】上記のイディオムの場合、イディオム見出し先頭の“as”と入力単語が一致しているので、イディオム見出しの次の単語“*a”に処理が移る。（ステップＳ１２２～Ｓ１２４、Ｓ１３０～Ｓ１３３）。“*a”は可変部分を表す代表記号なので、さらに次の単語“as”に処理が移る（ステップＳ１２２、Ｓ１３２、Ｓ１３３）。ここで、“as”は変化形展開指定を含まないので文字列比較を行い（ステップＳ１２４）、一致していることがわかる。

【００６７】次に、イディオムと入力文のそれぞれ次の単語、*x (can)と“could”のマッチングに処理が移る（ステップＳ１３０～Ｓ１３３）。*x (can) は変化形展開指定なのでステップＳ１２５へ進み、文字列の比較に入る前に、“can ”の助動詞としての活用変化を行う。

【００６８】活用変化は時制や単複などの語尾変化の他に、助動詞や“not”の付加も考慮するので、形態素解析が持つ語尾処理テーブルだけでなく、図１３に示す変化形テーブルを使って変化させる。本例では、“can”を変化展開した“could” “can”に入力の“can”が含まれることがわかる（ステップＳ１３６）ので、次の単語のマッチングを調べにいく（ステップＳ１３０～Ｓ１３３）。

【００６９】イディオムと入力文のそれぞれ、次の入力単語（“be”）とbeのマッチングに処理が移る。これも同様に一致していることがわかり、結局、固定部分のマッチングは成功することがわかる（ステップＳ１３３において成功終了）。

【００７０】次に、図１０の可変部分のマッチング処理（ステップＳ１１５）に入る。図１２に、この可変部分のマッチング処理のフローチャートを示す。最初に、可変部分の辞書引きを行う（ステップＳ１４１）。固定部分のマッチングの際に、代表記号＊aの対象単語が固定部分に挟まれたhappy でありえることがわかっているので、happy の辞書引きを行なう。

【００７１】次に、代表記号の中にユーザ代表記号があるかどうか調べ（ステップＳ１４２）、もしあればユーザ代表記号を定義本体部に置換する（ステップＳ１４３）。次に、形態素レベルでチェックできるかどうかを

16

調べる（ステップＳ１４４）。“*a”はシステム定義の単語品詞なので、“happy”の品詞が形容詞であることを確認し（ステップＳ１４５）、可変部分のマッチングが終了する。

【００７２】もし可変部分の対象単語が複数の単語からなる場合、すなわち代表記号が句品詞の場合には、構文処理が呼ばれ、指定の属性がチェックされる（ステップＳ１４６）。以上で、入力文がイディオム

英単語　[as *a as *x(can) be.]

訳語　この上なく *a で

を含むことがわかったので、最後に、次のようなイディオム部分の訳文を生成し、訳バッファに格納しておき、イディオム処理が完了する（ステップＳ１１６）。

イディオム部分 [as happy as could be]

訳文　　この上なく幸福で

【００７３】さらに処理を繰り返し（ステップＳ１１７、Ｓ１１８）、イディオムの次の単語から辞書引きを再開する。この例はイディオムの範囲が文末までなのでこの時点で辞書引きが完了する。以降、イディオム以外の“I ¥was”の単語列に関して、通常の構文解析、構文変換が行われ、最後に、翻訳文生成処理でイディオム以外の日本語訳「私は～あった」とイディオム部分の日本語訳「この上なく幸福で」の合成が行なわれ、次のような文全体の訳文が得られる。

「私は、この上なく幸福であった。」

【００７４】次に、利用者が登録した代表記号を用いたイディオムの例と、そのイディオムを用いた文の翻訳処理について説明する。利用者が新しい代表記号を登録するために、次のような書式を利用するものとする。

（新記号）“::=”（定義本体部）

ここでこのような書式で記述された代表記号は、ＣＰＵ３１によって辞書メモリ３６に格納される。

【００７５】また、定義本体部は次のように記述する。

“｛”（文法属性）“/”（値）“，”（文法属性）“/”（値）“，”……“｝”

指定できる文法属性と値（属性値）として、利用者には、翻訳システムが内部的に定めているあらゆるパラメータを開放する。これにより、文法や意味の詳細な制約を使って代表記号を定義でき、簡単にイディオム登録に利用できるようになる。属性と値は、例えば、図１４のような属性を考えることができる。

【００７６】このように形態素解析以外の様々のレベルの属性と値も利用者が指定できるようになると、辞書引きの段階だけでは、ある単語列が利用者が定義した代表記号かどうかのチェックができなくなる。例えば、ある単語列の品詞が「名詞句」であるかどうかは構文解析の段階まで進まないとわからない。図１４の各属性の値にはそれぞれどの段階でチェックできるかが明記されている。

【００７７】以下では、

15

"as　形容詞　a 冠詞無の単数名詞句 as … can"
「この上なく～」
というイディオムに基づき説明する。このイディオムの語順は特殊であるため、「冠詞無の単数名詞句」という文法的制約の指定が必要になる。
【００７８】以下に、この文法的制約を表す代表記号の登録処理と翻訳処理について説明を行なう。図１４の表を用いると、品詞(cat)が名詞句(n)で、活用形(inf)が単数(sg)、用法(use)が冠詞なし(detnil)という「冠詞無の単数名詞句」を表す代表記号は、
*Nsg ::= (cat/n, inf/sg, use/detnil)
と登録できる。
【００７９】この代表記号を用いることで、上記イディオムは次のように簡単に登録できる。
英単語　[as *a a *Nsg as *x(can) be]
訳語　この上なく*a *Nsg
【００８０】次に、このイディオムを用いた翻訳処理を図１２を用いて説明する。
[I bought as large a hat as could be.]
があったとすると，まずイディオムの検索から。固定部分のマッチングまでの処理が、上記実施例と同様に行なわれる。
【００８１】次に、可変部分のマッチングに移る。例では、このイディオムの中に利用者が定義した代表記号を含んでいるので（ステップＳ１４２），利用者代表記号を定義本体部に置換する処理を行なう（ステップＳ１４３）。
【００８２】本例の代表記号*Nsgの定義本体部の場合には、
(cat/n, inf/sg, use/detnil)
のように展開される。図１４を参照することで、このうち、cat/n, inf/sg は形態素解析レベルでチェックできる（ステップＳ１４５）のに対して，use/detnilは構文解析まで進んで初めて，冠詞無名詞であることがわかる（ステップＳ１４６）。すなわち本例では、"a hat" の "a" はイディオム見出し中に含まれるので、可変部分は "hat" だけになり、冠詞無名詞(*Nsg)であることがわかる。
【００８３】上記のようにマッチング処理が成功し，イディオムを認識した後は、上記実施例と同様に処理が進み、イディオム部分の訳「この上なく大きな帽子」が得られる。
イディオム部分 [as large a hat as could be]
訳文　この上なく大きな帽子
さらに，文全体の訳文
「私はこの上なく大きな帽子を買った」
が得られる。
【００８４】
【発明の効果】この発明によれば、イディオムの見出し語の固定部分を変化展開記号を含む表現で登録し、かつ

16

変化展開記号で表現された固定部分を変形表現に生成展開し、この変形表現と変化展開記号が存在する位置に相当する単語又は単語列との同定を行うため、固定部分が種々の変形表現されることのある複数個のイディオムの見出し語を、１つの見出し語で登録することができる。よって、固定部分の種々の変形表現について考えられるあらゆるパターンを登録する必要はなく，登録されるイディオムの見出し語の増大を押さえることができ、さらにイディオム記憶容量及び検索時間を抑制できる。
【００８５】また，この発明によれば、新たに定義された属性及びその属性値を有する単語又は単語列を代表する第２の代表記号を用いてイディオムの見出し語の可変部分を表現し、かつイディオムの同定を行うので，イディオムとして登録する見出し語の長さを揃えることができると共に、利用者自身にとって必要な、あるいは、ある分野の文章に特有な表現形式を持つイディオムの登録及び同定をすることができる。
【図面の簡単な説明】
【図１】この発明の機械翻訳装置の基本構成を示すブロック図である。
【図２】機械翻訳の概念を模式的に示す図である。
【図３】この発明の一実施例の構成ブロック図である。
【図４】図３に示される翻訳モジュール３５のブロック図である。
【図５】バッファＡの格納内容を模式的に示す図である。
【図６】バッファＢの格納内容を模式的に示す図である。
【図７】バッファＣの格納内容を模式的に示す図である。
【図８】バッファＤの格納内容を模式的に示す図である。
【図９】バッファＥの格納内容を模式的に示す図である。
【図１０】辞書引き・イディオム処理を示すフローチャートである。
【図１１】固定部分のマッチング処理を示すフローチャートである。
【図１２】可変部分のマッチング処理を示すフローチャートである。
【図１３】変化形テーブルを模式的に示す図である。
【図１４】代表記号として指定できる属性を模式的に示す図である。
【符号の説明】
１　入力手段
２　イディオム登録手段
３　辞書引き・形態素解析手段
３ａ　品詞抽出部
３ｂ　訳語抽出部
４　イディオム翻訳手段

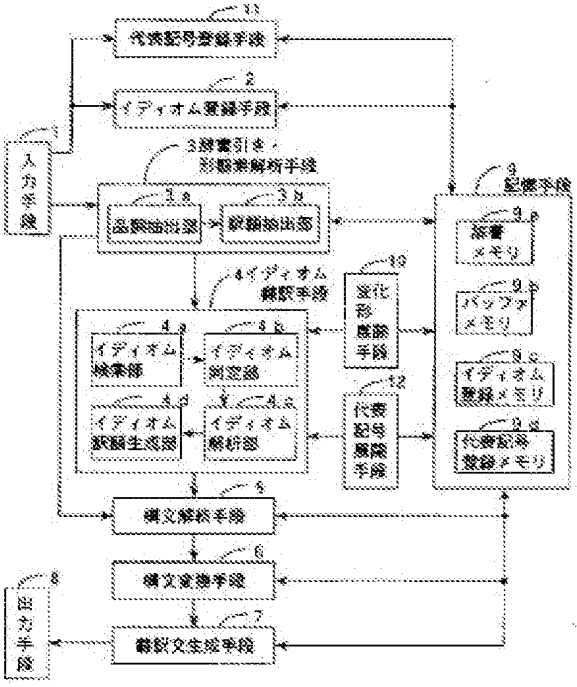
17

4a　イディオム検索部
4b　イディオム同定部
4c　イディオム解析部
4d　イディオム訳語生成部
5　構文解析手段
6　構文変換手段
7　翻訳文生成手段
8　出力手段
9　記憶手段
9a　辞書メモリ
9b　バッファメモリ
9c　イディオム登録メモリ

18

9d　代表記号登録メモリ
10　変化形展開手段
11　代表記号登録手段
12　代表記号展開手段
31　メインCPU
32　メインメモリ
33　表示装置
34　キーボード
35　翻訳モジュール
36　辞書メモリ
37　バス

【図1】



【図3】

例文　（This is a pen.）

バッファA　：　原文バッファ

t	h	i	s				
i	s						
a							
p	e	n					
.							

【図6】

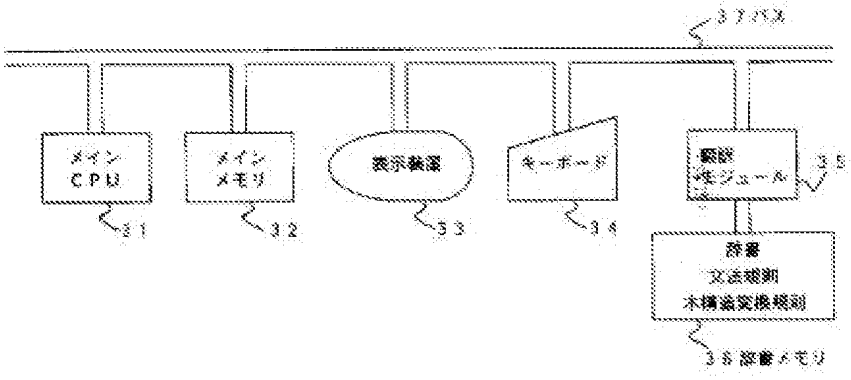
バッファB　：　辞書引き結果のバッファの一部

this	代名詞	指示形容詞	
is	動詞		
a	冠詞		
pen	名詞		

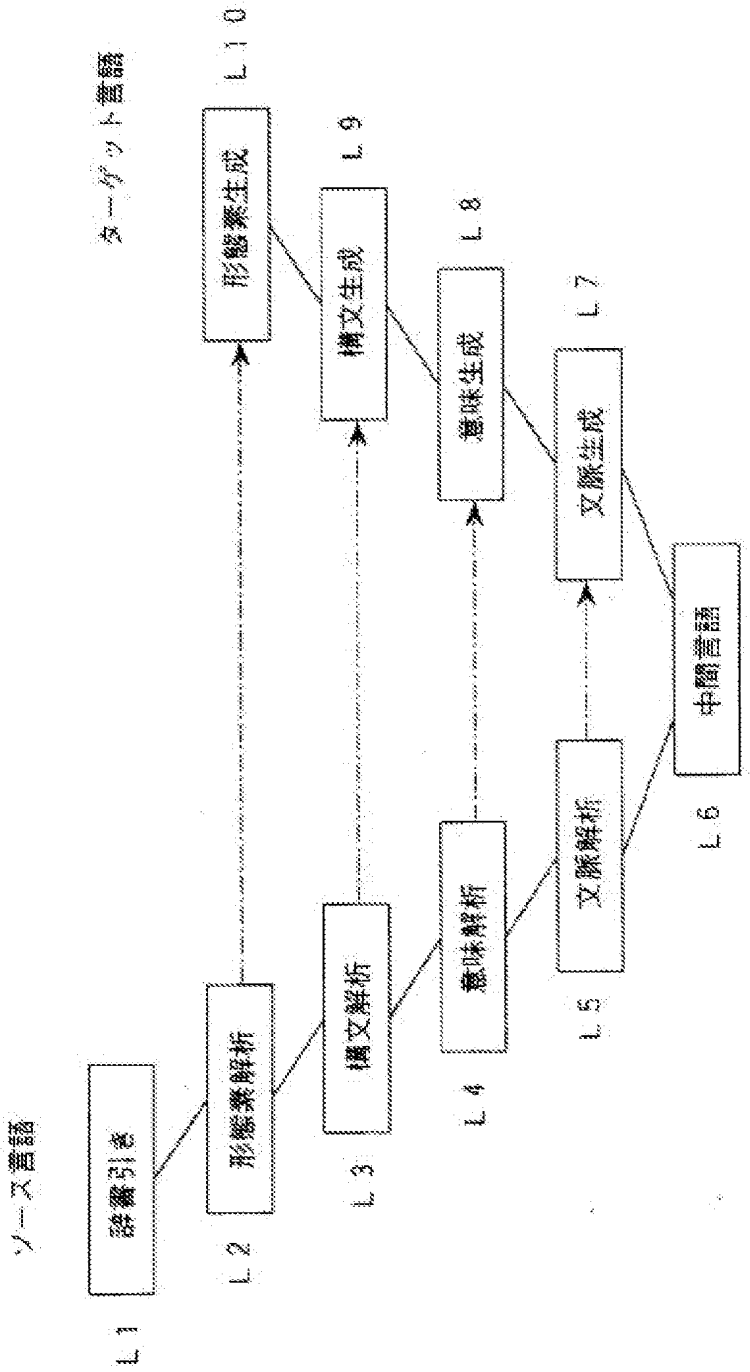
【図9】

バッファC　：　出力文バッファ

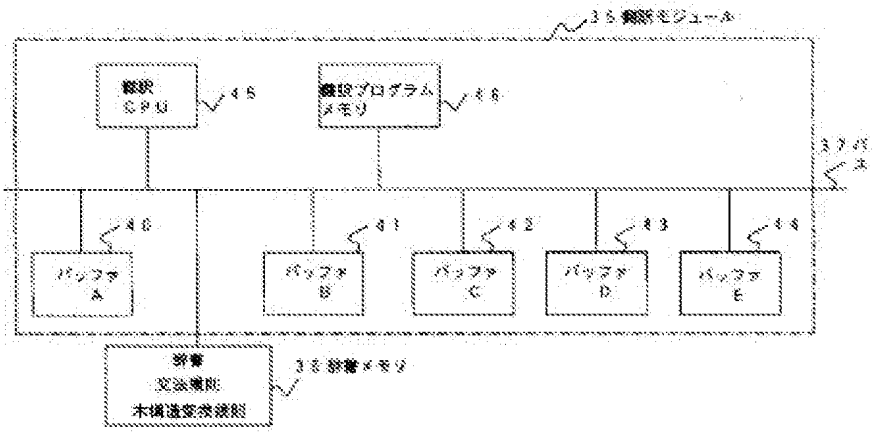
これはペンである。



【図2】

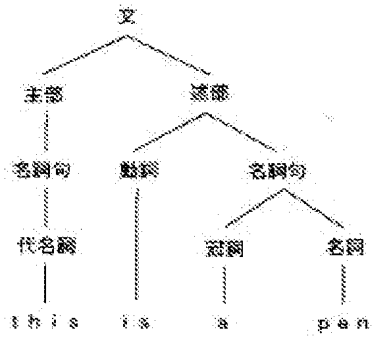

ソース言語
ターゲット言語
辞書引き
形態素解析
構文解析
意味解析
文脈解析
中間言語
形態素生成
構文生成
意味生成
文脈生成
L1
L2
L3
L4
L5
L6
L7
L8
L9
L10

【図4】



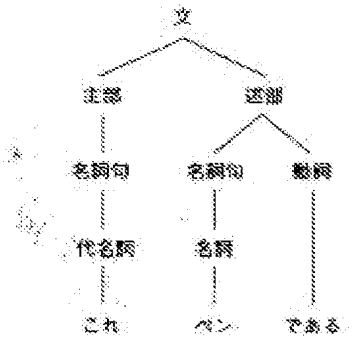
【図7】

バッファC　：　構文解析結果のバッファ



【図8】

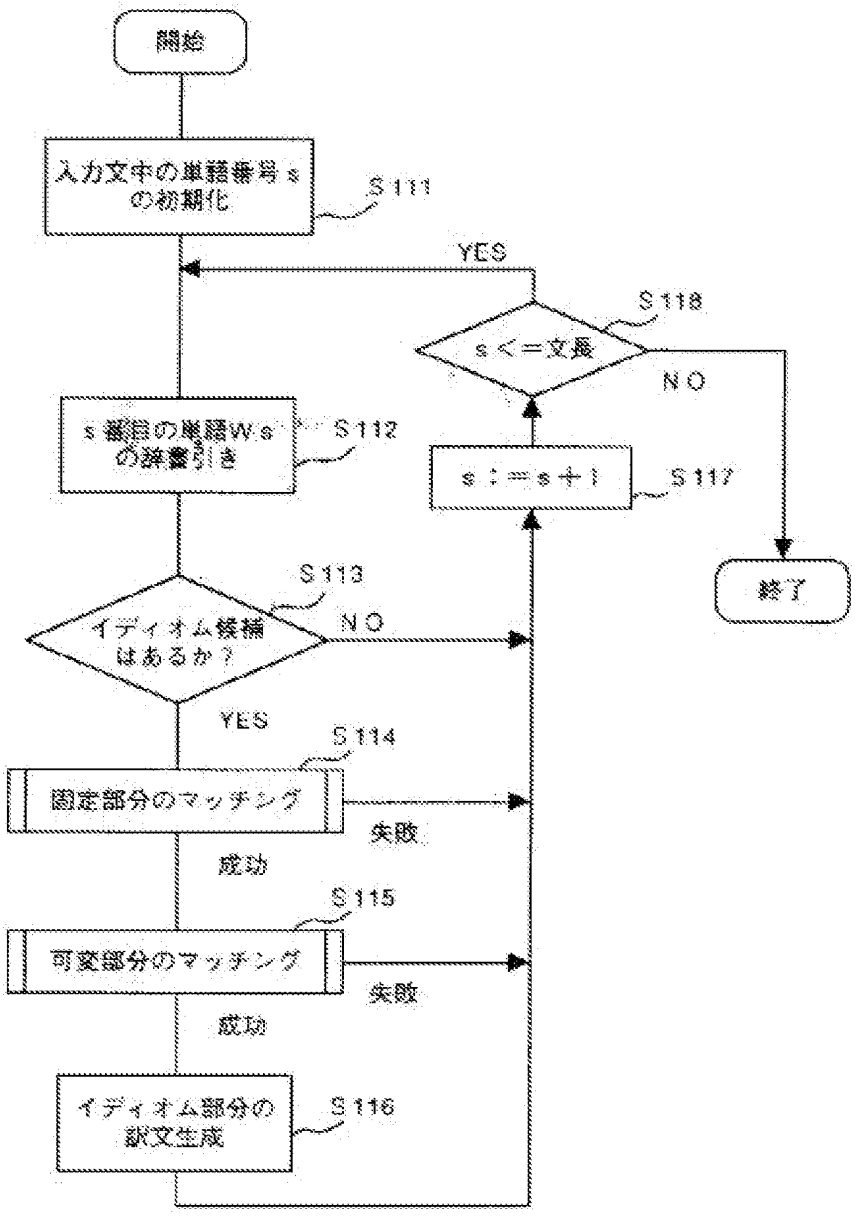
バッファD　：　木変換した結果のバッファ



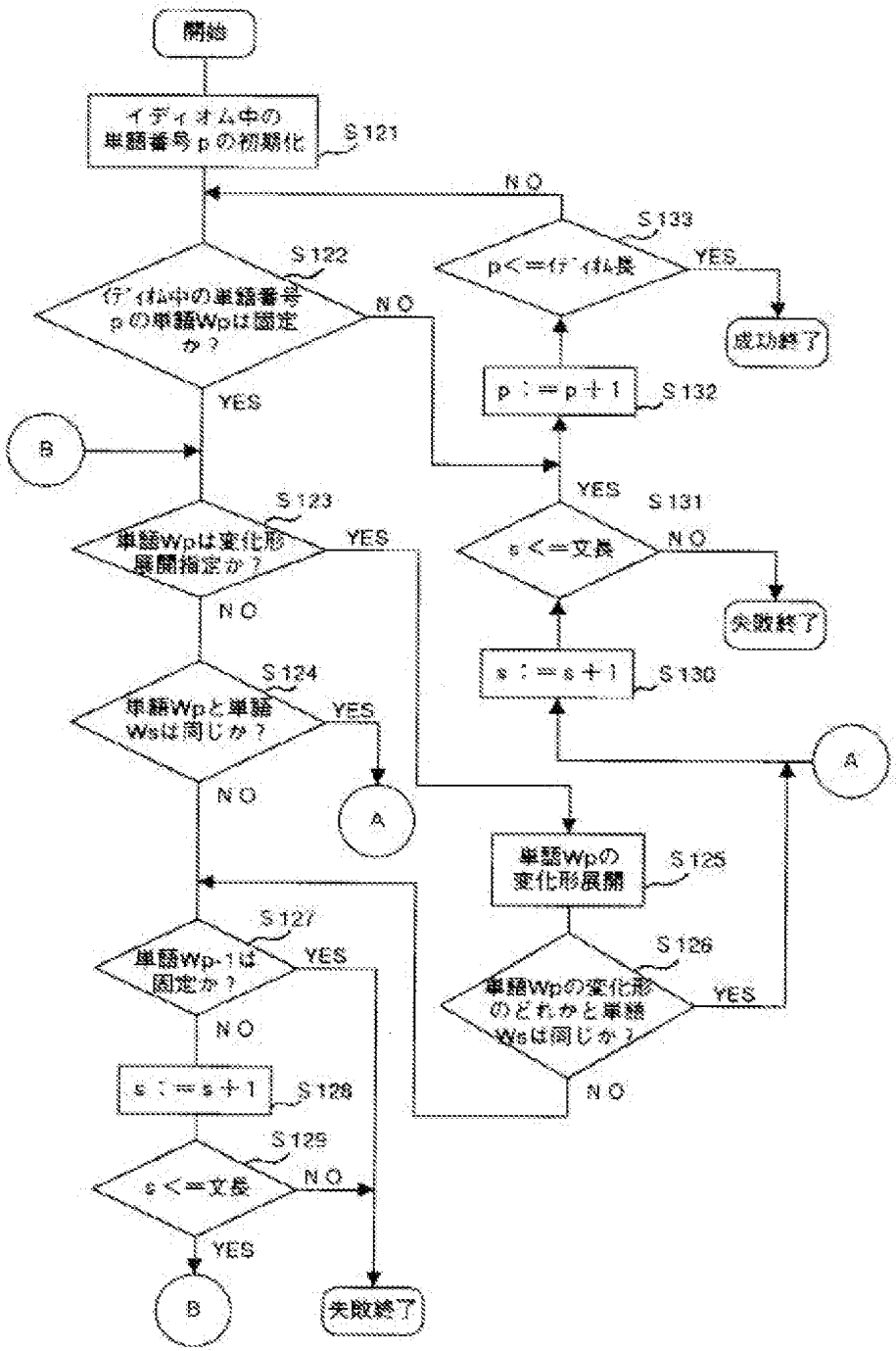
【図13】

単形	品詞	変化形
can	x	can , could
	v	can , cans , canned
	n	can , cans
be	v	is , was , were , will be , is not , was not , were not , will not be
hat	n	hat , hats

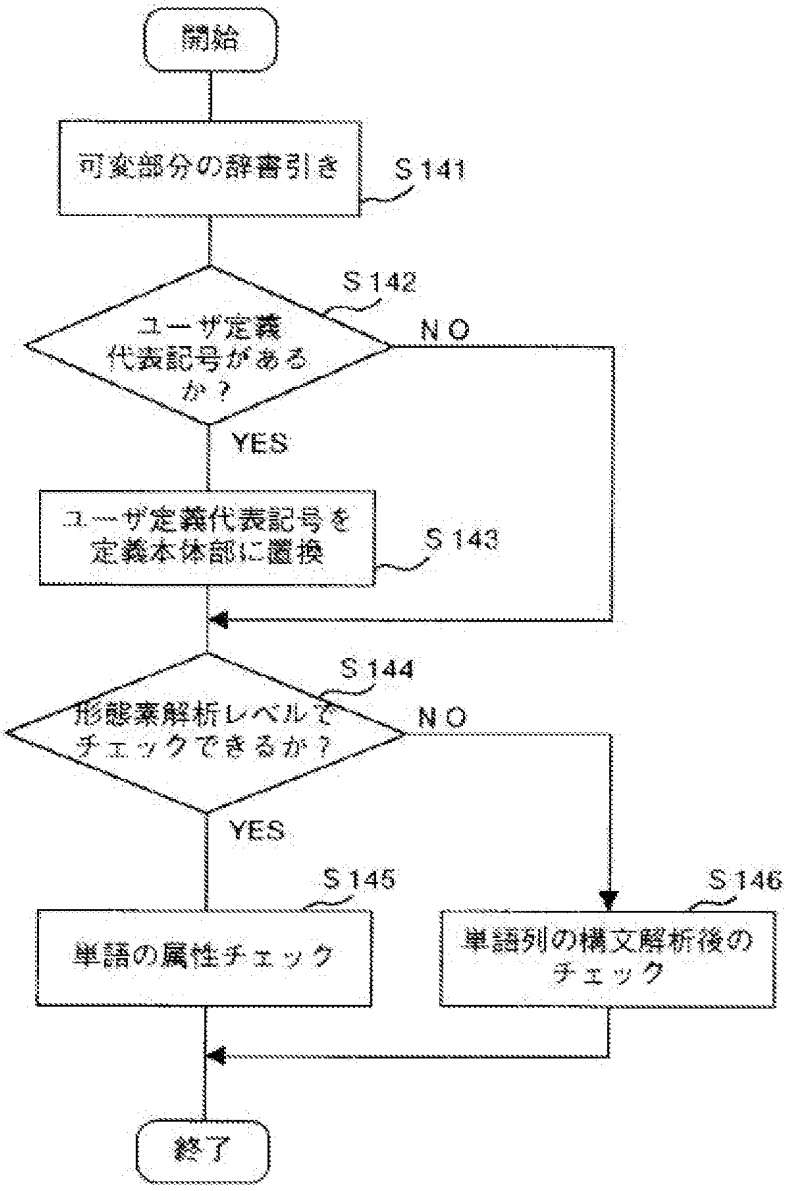
【図10】


開始
入力文中の単語番号sの初期化
S111
YES
S118
s<=文長
NO
s番目の単語Wsの辞書引き
S112
s:=s+1
S117
終了
S113
イディオム候補はあるか？
NO
YES
S114
固定部分のマッチング
失敗
成功
S115
可変部分のマッチング
失敗
成功
イディオム部分の訳文生成
S116

【図11】


開始
イディオム中の単語番号pの初期化
S121
S122
イディオム中の単語番号pの単語Wpは固定か？
NO
YES
S123
単語Wpは変化形展開指定か？
YES
NO
S124
単語Wpと単語Wsは同じか？
YES
NO
A
B
S125
単語Wpの変化形展開
S126
単語Wpの変化形のどれかと単語Wsは同じか？
YES
NO
S127
単語Wp-1は固定か？
YES
NO
s：＝s＋1
S128
S129
s＜＝文長
NO
YES
B
失敗終了
S131
s＜＝文長
NO
YES
失敗終了
s：＝s＋1
S130
A
S133
p＜＝イディオム長
YES
NO
成功終了
p：＝p＋1
S132

【図12】


開始
可変部分の辞書引き
S141
S142
ユーザ定義
代表記号がある
か？
NO
YES
ユーザ定義代表記号を
定義本体部に置換
S143
S144
形態素解析レベルで
チェックできるか？
NO
YES
S145
単語の属性チェック
S146
単語列の構文解析後の
チェック
終了

【図１４】

属性	属性値	意味	チェック処理
品詞cat	n	名詞	形態素解析
	v	動詞	形態素解析
	a	形容詞	形態素解析
	d	副詞	形態素解析
	np	名詞句	構文解析
	vp	動詞句	構文解析
	…		
変化形inf	ing	進行形	形態素解析
	ed	過去形	形態素解析
	en	過去分詞形	形態素解析
	pl	複数	形態素解析
	sg	単数	形態素解析
	er	比較級	形態素解析
	est	最大級	形態素解析
	…		
意味sem	hum	人間	意味解析
	anim	可動物	意味解析
	prod	生産物	意味解析
	nat	自然	意味解析
	mobj	精神的物体	意味解析
	ploc	物理的場所	意味解析
	time	時間	意味解析
	…		
用法use	deta	冠詞aがつく	構文解析
	detthe	冠詞theがつく	構文解析
	detnil	冠詞なし	構文解析
	…		

フロントページの続き

(72)発明者　九津見　毅
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内